November 2024

A new golden age of discovery

Seizing the *AI for Science* opportunity

Conor Griffin | Don Wallace | Juan Mateos-Garcia | Hanna Schieve | Pushmeet Kohli

Acknowledgements

Thank you to Louisa Bartolo, Zoë Brammer and Nick Swanson for research support, and to the following individuals who shared insights with us through interviews and/or feedback on the draft. All views, and any mistakes, belong solely to the authors.

Žiga Avsec, Nicklas Lundblad, John Jumper, Matt Clifford, Ben Southwood, Craig Donner, Joëlle Barral, Tom Zahavy, Been Kim, Sebastian Nowozin, Matt Clancy, Matej Balog, Jennifer Beroshi, Nitarshan Rajkumar, Brendan Tracey, Yannis Assael, Massimiliano Ciaramita, Michael Webb, Agnieszka Grabska-Barwinska, Alessandro Pau, Tom Lue, Agata Laydon, Anna Koivuniemi, Abhishek Nagaraj, Harry Law, Tom Westgarth, Guy Ward-Jackson, Arianna Manzini, Stefano Bianchini, Sameer Velankar, Ankur Vora, Sébastien Krier, Joel Z Leibo, Elisa Lai H. Wong, Ben Johnson, David Osimo, Andrea Huber, Dipanjan Das, Ekin Dogus Cubuk, Jacklynn Stott, Kelvin Guu, Kiran Vodrahalli, Sanil Jain, Trieu Trinh, Rebeca Santamaria-Fernandez, Remi Lam, Victor Martin, Neel Nanda, Nenad Tomasev, Obum Ekeke, Uchechi Okereke, Francesca Pietra, Rishabh Agarwal, Peter Battaglia, Anil Doshi, Yian Yin, Sukhdeep Singh. ■

Introduction

Introduction

A quiet revolution is brewing in labs around the world, where scientists' use of AI is growing exponentially. One in three postdocs now use large language models to help carry out literature reviews, coding, and editing. In October, the creators of our AlphaFold 2 system, Demis Hassabis and John Jumper became Nobel Laureates in Chemistry for using AI to predict the structure of proteins, alongside the scientist David Baker, for his work to design new proteins. Society will soon start to feel these benefits more directly, with drugs and materials designed with the help of AI currently making their way through development.

In this essay, we take a tour of how AI is transforming scientific disciplines from genomics to computer science to weather forecasting. Some scientists are training their own AI models, while others are fine-tuning existing AI models, or using these models' predictions to accelerate their research. Scientists are using AI as a scientific instrument to help tackle important problems, such as designing proteins that bind more tightly to disease targets, but are also gradually transforming how science itself is practised.

There is a growing imperative behind scientists' embrace of AI. In recent decades, scientists have continued to deliver consequential advances, from Covid-19 vaccines to renewable energy. But it takes an ever larger number of researchers to make these breakthroughs, and to transform them into downstream applications. As a result, even though the scientific workforce has grown significantly over the past half-century, rising more than seven fold in the US alone, the societal progress that we would expect to follow, has slowed. For instance, much of the world has witnessed a sustained slowdown in productivity growth that is undermining the quality of public services. Progress towards the 2030 Sustainable Development Goals, which capture the biggest challenges in health, the environment, and beyond, is stalling.

In particular, scientists looking to make breakthroughs today increasingly run into challenges relating to *scale* and *complexity*, from the ever-growing literature base they need to master, to the increasingly complex experiments they want to run. Modern deep learning methods are particularly well-suited to these scale and complexity challenges and can compress the time that future scientific progress would otherwise require. For instance, in structural biology, a single x-ray crystallography experiment to determine the structure of a protein can take years of work and cost approximately $100,000, depending on the protein. The AlphaFold Protein Structure Database now provides instant access to 200 million predicted protein structures for free.

The potential benefits of AI to science are not guaranteed. A significant share of scientists already use LLM-based tools to assist with everyday tasks, such as coding and editing, but the share of scientists using AI-centric research approaches is much lower, albeit rising rapidly. In the rush to use AI, some early scientific use cases have had questionable impact. Policymakers can help accelerate AI's use and steer it towards higher-impact areas. The US Department of Energy, the European Commission, the UK's Royal Society, and the US National Academies, among others, have recently recognised the *AI for Science* opportunity. But no country has yet put a comprehensive strategy in place to enable it.

We hope our essay can inform such a strategy. It is aimed at those who make and influence science policy, and funding decisions. We first identify 5 opportunities where there is a growing imperative to use AI in science and examine the primary ingredients needed to make breakthroughs in these areas. We then explore the most commonly-cited risks from using AI in science, such as to scientific creativity and reliability, and argue that AI can ultimately be net beneficial in each area. We conclude with four public policy ideas to help usher in a new golden age of AI-enabled science.

Throughout the essay we draw on insights from over two dozen interviews with experts from our own AI for Science projects, as well as external experts. The essay naturally reflects our vantage point as a private sector lab, but we believe the case we make is relevant for the whole of science. We hope that readers will respond by sharing their take on the most important AI for Science opportunities, ingredients, risks and policy ideas. ■

Part A: The opportunities 06
Part B: The ingredients 13
Part C: The risks 25
Part D: The policy response 32

Part A

The opportunities

The opportunities

Scientists aim to understand, predict, and influence how the natural and social worlds work, to inspire and satisfy curiosity, and to tackle important problems facing society. Technologies and methods, like the microscope, x-ray diffraction, and statistics, are both *products* of science and *enablers* of it. Over the past century, scientists have increasingly relied on these instruments to carry out their experiments and advance their theories. Computational tools and large-scale data analysis have become particularly important, enabling everything from the discovery of the Higgs boson to the mapping of the human genome. From one view, scientists' growing use of AI is a logical extension of this long-running trend. But it may also signal something much more profound - a discontinuous leap in the limits of what science is capable of.

Rather than listing all areas where it is *possible* to use AI, we highlight five opportunities where we think there is an *imperative* to use it. These opportunities apply across disciplines and address a specific bottleneck, related to scale and complexity, that scientists increasingly face at different points in the scientific process, from generating powerful novel hypotheses to sharing their work with the world. ■

5 opportunities to accelerate science with AI

1. Knowledge
Transform how scientists digest and communicate knowledge

2. Data
Generate, extract, and annotate large scientific datasets

3. Experiments
Simulate, accelerate and inform complex experiments

4. Models
Model complex systems and how their components interact

5. Solutions
Identify novel solutions to problems with large search spaces

1. Knowledge

Transform how scientists digest and communicate knowledge

To make new discoveries, scientists need to master a pre-existing body of knowledge that continues to grow exponentially and become ever more specialised. This 'burden of knowledge' helps explain why scientists making transformative discoveries are increasingly older, interdisciplinary, and located at elite universities, and why the share of papers authored by individuals, or small teams, is declining, even though small teams are often better-placed to advance disruptive scientific ideas. When it comes to sharing their research there have been welcome innovations such as preprint servers and code repositories, but most scientists still share their findings in dense, jargon-heavy, English-only papers. This can impede rather than ignite interest in scientists' work, including from policymakers, businesses, and the public.

Scientists are already using LLMs, and early scientific assistants based on LLMs, to help address these challenges, such as by synthesising the most relevant insights from the literature. In an early demonstration, our Science team used our Gemini LLM to find, extract, and populate specific data from the most relevant subset of 200,000 papers, within a day. Upcoming innovations, such as fine-tuning LLMs on more scientific data and advances in long context windows and citation use, will steadily improve these capabilities. As we expand on below, these opportunities are not without risk. But they provide a window to fundamentally rethink certain scientific tasks, such as what it means to 'read' or 'write' a scientific paper in a world where a scientist can use an LLM to help critique it, tailor its implications for different audiences, or transform it into an 'interactive paper' or audio guide. ■

---

2. Data

Generate, extract, and annotate large scientific datasets

Despite popular narratives about an era of data abundance, there is a chronic lack of scientific data on most of the natural and social world, from the soil, deep ocean and atmosphere, to the informal economy. AI could help in different ways. It could make existing data collection more accurate, for example by reducing the noise and errors that can occur when sequencing DNA, detecting cell types in a sample, or capturing animal sounds. Scientists can also exploit LLMs' growing ability to operate across images, video and audio, to extract the *unstructured* scientific data that is buried in scientific publications, archives, and less obvious resources such as instructional videos, and convert it into structured datasets.

AI can also help to annotate scientific data with the supporting information that scientists need in order to use it. For example, at least one-third of microbial proteins are not reliably annotated with details about the function(s) that they are thought to perform. In 2022, our researchers used AI to predict the function of proteins, leading to new entries in the UniProt, Pfam and InterPro databases.

AI models, once validated, can also serve as new sources of synthetic scientific data. For example, our AlphaProteo protein design model is trained on more than 100 million AI-generated protein structures from AlphaFold 2, along with experimental structures from the Protein Data Bank. These AI opportunities can complement and increase the return on other much-needed efforts to generate scientific data, such as digitising archives, or funding new data capture technologies and methods, like efforts underway in single cell genomics to create powerful datasets of individual cells in unprecedented detail. ■

3. Experiments

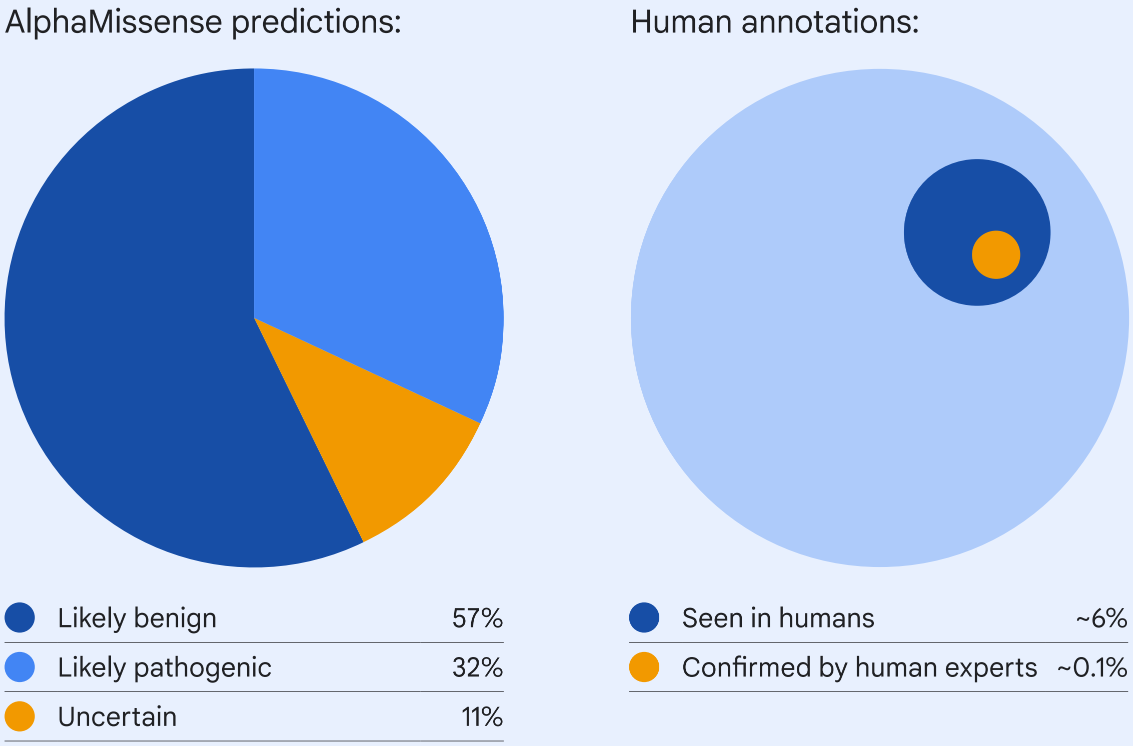
Simulate, accelerate and inform complex experiments

Many scientific experiments are expensive, complex, and slow. Some do not happen at all because researchers cannot access the facilities, participants or inputs that they need. Fusion is a case in point. It promises an energy source that is practically limitless, emission-free and could enable the scaling of energy-intensive innovations, like desalination. To realise fusion, scientists need to create and control *plasma* - a fourth fundamental state of matter. However, the facilities needed are hugely complex to build. ITER's prototype tokamak reactor began construction in 2013, but plasma experiments are not set to begin until the mid-2030s at the earliest, although others hope to build smaller reactors on shorter timelines.

AI could help to simulate fusion experiments and enable much more efficient use of subsequent experiment time. One approach is to run reinforcement learning agents on simulations of physical systems. Between 2019 and 2021, our researchers partnered with the Swiss Federal Institute of Technology Lausanne to demonstrate how to use RL to control the shape of plasma in a simulation of a tokamak reactor. These approaches could be extended to other experimental facilities, such as particle accelerators, telescope arrays, or gravitational wave detectors.

Using AI to simulate experiments will look very different across disciplines, but a common thread is that the simulations will often inform and guide physical experiments, rather than substitute for them. For example, the average person has more than 9,000 *missense* variants, or single letter substitutions in their DNA. Most of these genetic variants are benign but some can disrupt the functions that proteins perform, contributing to rare genetic diseases like cystic fibrosis as well as common diseases like cancer. Physical experiments to test the effects of these variants are often limited to a single protein. Our AlphaMissense model classifies 89% of the 71 million potential human missense variants as likely harmful or benign, enabling scientists to focus their physical experiments on the most likely contributors to disease. ■

All possible 71 million human missense variants



AlphaMissense predicted the pathogenicity of all possible 71 million missense variants. It classified 89% - predicting 57% were likely benign and 32% were likely pathogenic.

4. Models

Model complex systems and how their components interact

In a 1960 paper, the Nobel Prize winning physicist Eugene Wigner marvelled at the “unreasonable effectiveness” of mathematical equations for modelling important natural phenomena, such as planetary motion. However, over the past half century, models that rely on sets of equations or other deterministic assumptions have struggled to capture the full complexity of systems in biology, economics, weather, and elsewhere. This reflects the sheer number of interacting parts that make up these systems, as well as their dynamism and potential for emergent, random or chaotic behaviour. The challenges in modelling these systems impedes scientists’ ability to predict or control how they will behave, including during shocks or interventions, such as rising temperatures, a new drug, or the introduction of a tax change.

AI could more accurately model these complex systems by ingesting more data about them, and learning more powerful patterns and regularities within this data. For example, modern weather forecasting is a triumph of science and engineering. For governments and industry, it informs everything from renewable energy planning to preparing for hurricanes and floods. For the public, the weather is the most popular non-branded query on Google Search. Traditional numeral prediction methods are based on carefully-defined physics equations that provide a very useful, yet imperfect, approximation of the atmosphere's complex dynamics. They are also computationally expensive to run. In 2023, we released a deep learning system that predicts weather conditions up to 10 days in advance, which outperformed traditional models on accuracy and prediction speed. As we expand on below, using AI to forecast weather variables could also help to mitigate and respond to climate change. For instance, when pilots fly through humid regions it can cause *condensation trails* that contribute to aviation's global warming impact. Google scientists recently used AI to predict when and where humid regions may arise to help pilots avoid flying through them.

In many cases, AI will enrich traditional approaches to modelling complex systems rather than replace them. For example, agent-based modelling simulates interactions between individual actors, like firms and consumers, to understand how these interactions might affect a larger more complex system like the economy. Traditional approaches require scientists to specify beforehand how these computational agents should behave. Our research teams recently outlined how scientists could use LLMs to create more flexible *generative* agents that communicate and take actions, such as searching for information or making purchases, while also reasoning about and remembering these actions. Scientists could also use reinforcement learning to study how these agents learn and adapt their behaviour in more dynamic simulations, for example in response to the introduction of new energy prices or pandemic response policies. ■

---

5. Solutions

Identify novel solutions to problems with large search spaces

Many important scientific problems come with a practically incomprehensible number of potential solutions. For example, biologists and chemists aim to determine the structure, characteristics, and function(s) of molecules such as proteins. One goal of such work is to help design novel versions of these molecules to serve as antibody drugs, plastic-degrading enzymes or new materials. However, to design a small molecule drug, scientists face more than 10^{60} potential options. To design a protein with 400 standard amino acids, they face 20^{400} options. These large search spaces are not limited to molecules but are commonplace for many scientific problems, such as finding the best proof for a maths problem, the most efficient algorithm for a computer science task, or the best architecture for a computer chip.

Traditionally, scientists rely on some combination of intuition, trial and error, iteration, or brute force computing to find the best molecule, proof, or algorithm. However, these methods struggle to exploit the huge space of potential solutions, leaving better ones undiscovered. AI can open up new parts of these search spaces while also homing in more quickly on the solutions that are most likely to be viable and useful - a delicate balancing act. For example, in July, our AlphaProof and AlphaGeometry 2 systems correctly solved four out of six problems from the International Mathematical Olympiad, an elite high school competition. The systems make use of our Gemini LLM architecture to generate a large number of novel ideas and potential solutions for a given maths problem, and combine this with systems grounded in mathematical logic that can iteratively work towards the candidate solutions that are most likely to be correct. ■

AI scientists or AI-empowered scientists?

This growing use of AI in science, and the emergence of early AI scientific assistants, raises questions about how fast and how far the capabilities of AI may advance and what this will mean for human scientists. Current LLM-based AI scientific assistants make a relatively small contribution to a relatively narrow range of tasks, such as supporting literature reviews. There are plausible near-term scenarios in which they become better at these tasks and become capable of more impactful ones, such as helping to generate powerful hypotheses, or helping to predict the outcomes of experiments. However, current systems still struggle with the deeper creativity and reasoning that human scientists rely on for such tasks. Efforts are underway to improve these AI capabilities, for example by combining LLMs with logical deduction engines, as in our AlphaProof and AlphaGeometry 2 examples, but further breakthroughs are needed. The ability to accelerate or automate experiments will also be harder for those that require complicated actions in wet labs, interacting with human participants, or lengthy processes, such as monitoring disease progression. Although again, work is underway in some of these areas, such as new types of laboratory robotics and automated labs.

Even as AI systems' capabilities improve, the greatest marginal benefit will come from deploying them in use cases that play to their relative strengths - such as the ability to rapidly extract information from huge datasets - and which help address genuine bottlenecks to scientific progress such as the five opportunities outlined above, rather than automating tasks that human scientists already do well. As AI enables cheaper and more powerful science, demand for science and scientists will also grow. For example, recent breakthroughs have already led to a slew of new startups in areas like protein design, material science and weather forecasting. Unlike other sectors, and despite past claims to the contrary, future demand for science appears practically limitless. New advances have always opened up new, unpredictable regions in the scientific *map of knowledge*, and AI will do similar. As envisioned by Herbert Simon, AI systems will also become *objects* of science research, with scientists set to play a leading role in evaluating and explaining their scientific capabilities, as well as in developing new types of human-AI scientific systems. ■

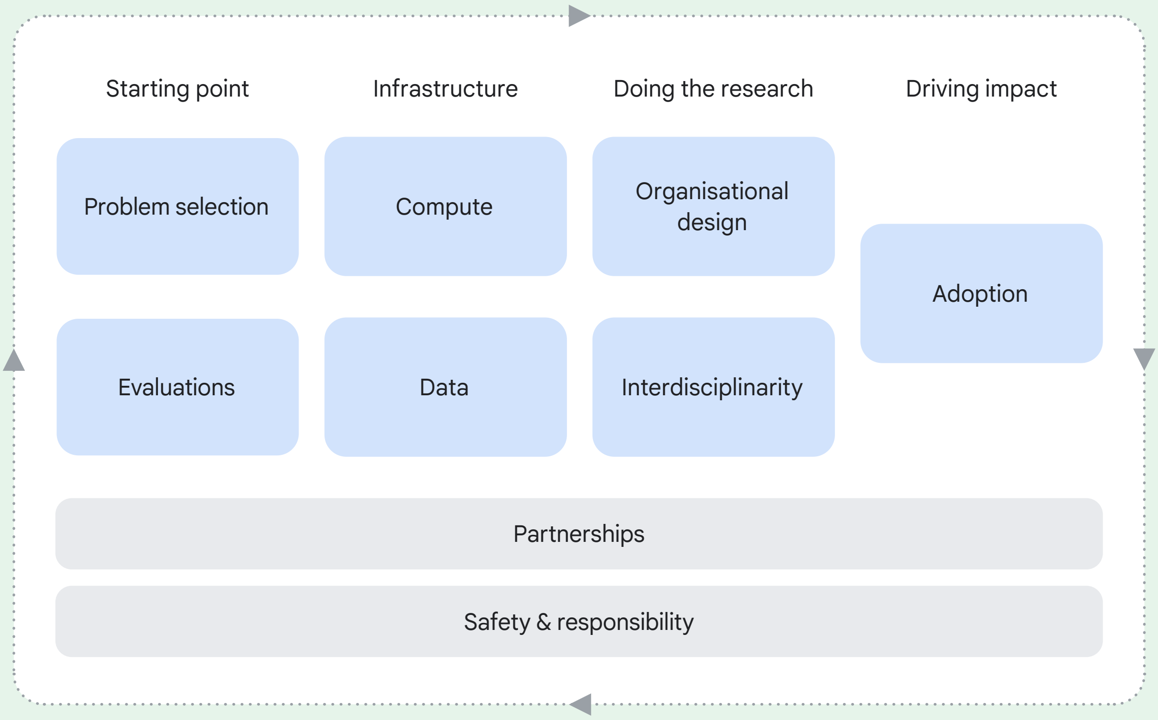
Part B

The ingredients

The ingredients

We are interested in the ingredients that ambitious AI for Science efforts need to succeed - both at the individual research effort level and at the level of the science ecosystem, where policymakers have more scope to shape them. The experts that we interviewed routinely cited several ingredients that we organised into a toy model, which we call *the AI for Science production function*. This production function is not meant to be exhaustive, prescriptive, or a neat linear process. The ingredients will be intuitive to many, but our interviews revealed a number of lessons about what they look like in practice which we share below. ■

The AI for science production function



1. Problem selection

Pursue ambitious, AI-shaped problems

Scientific progress rests on being able to identify an important problem and ask the *right* question about how to solve it. In their exploration into the genesis of scientific breakthroughs, Venkatesh Narayanamurti and Jeffrey Y. Tsao document how important the reciprocal and recursive relationship between questions and answers is, including the importance of asking ambitious new questions. Our Science team starts by thinking about whether a potential research problem is significant enough to justify a substantial investment of time and resources. Our CEO Demis Hassabis has a mental model to guide this assessment: thinking about all of science as a *tree of knowledge*. We are particularly interested in the roots - fundamental 'root node problems' like protein structure prediction or quantum chemistry that, if solved, could unlock entirely new branches of research and applications.

To assess whether AI will be suitable and additive, we look for problems with certain characteristics, such as huge combinatorial search spaces, large amounts of data, and a clear objective function to benchmark performance against. Often a problem is suitable for AI in principle, but the inputs aren't yet in place and it needs to be stored for later. One of the original inspirations for AlphaFold was conversations that Demis had many years prior as a student with a friend who was obsessed with the protein folding problem. Many recent breakthroughs also feature this coming together of an important scientific problem and an AI approach that has reached a point of maturity. For example, our fusion effort was aided by a novel reinforcement learning algorithm called *maximum a posteriori policy optimization*, which had only just been released. Alongside a new fast and accurate simulator that our partners EPFL had just developed, that enabled the team to overcome a data paucity challenge.

In addition to picking the right problem, it is important to *specify* it at the right level of difficulty. Our interviewees emphasised that a powerful problem statement for AI is often one that lends itself to *intermediate* results. If you pick a problem that's too hard then you won't generate enough signal to make progress. Getting this right relies on intuition and experimentation. ■

2. Evaluations

Invest in evaluation methods that can provide a robust performance signal and are endorsed by the community

Scientists use evaluation methods, such as benchmarks, metrics and competitions, to assess the scientific capabilities of an AI model. Done well, these evaluations provide a way to track progress, encourage innovation in methods, and galvanise researchers' interest in a scientific problem. Often, a variety of evaluation methods are required. For example, our weather forecasting team started with an initial 'progress metric' based on a few key variables, such as surface temperature, that they used to 'hill climb', or gradually improve their model's performance. When the model had reached a certain level of performance, they carried out a more comprehensive evaluation using more than 1,300 metrics inspired by the European Centre for Medium-Range Weather Forecasts's evaluation scorecard. In past work, the team learned that AI models can sometimes achieve good scores on these metrics in undesirable ways. For example, 'blurry' predictions - such as predicting rainfall within a large geographical area - are less penalised than 'sharp' predictions - such as predicting a storm in a location that is very slightly different to the actual location - the so-called 'double-penalty' problem. To provide further verification, the team evaluated the usefulness of their model on downstream tasks, such as its ability to predict the track of a cyclone, and to characterise the strength of 'atmospheric rivers' - narrow bands of concentrated moisture that can lead to flooding.

The most impactful AI for Science evaluation methods are often community-driven or endorsed. A gold standard is the Critical Assessment of protein Structure Prediction competition. Established in 1994 by Professor John Moult and Professor Krzysztof Fidelis, the biennial CASP competition has challenged research groups to test the accuracy of their protein structure prediction models against real, unreleased, experimental protein structures. It has also become a unique global community and a catalyst for research progress, albeit one that is hard to replicate quickly. The need for community buy-in also provides an argument for publishing benchmarks so that researchers can use, critique and improve them. However, this also creates the risk that the benchmark will 'leak' into an AI model's training data, reducing its usefulness for tracking progress. There is no perfect solution to this tradeoff but, at a minimum, new public benchmarks are needed at regular intervals. Scientists, AI labs and policymakers should also explore new ways to assess the scientific capabilities of AI models, such as setting up new third-party assessor organisations, competitions, and enabling more open-ended probing of AI models' capabilities by scientists. ■

3. Compute

Track how compute use is evolving and invest in specialist skills

Multiple government reviews have recognised the growing importance of compute to progress in AI, science, and the wider economy. As we expand on further below, there is also a growing focus on its energy consumption and greenhouse gas emissions. AI labs and policymakers should take a grounded, long-term view that considers how compute needs will vary across AI models and use cases, potential multiplier effects and efficiency gains, and how this compares to counterfactual approaches to scientific progress that don't use AI.

For example, some state-of-the-art AI models, such as in protein design, are relatively small. Larger models, like LLMs, are compute-intensive to train but typically require much less compute to fine-tune, or to run inference against, which can open up more efficient pathways to science research. Once an LLM is trained, it is also easier to make it more efficient, for example via better data curation, or by 'distilling' the large model into a smaller one. AI compute needs should also be evaluated in comparison to other models of scientific progress. For example, AI weather forecasting models are compute-intensive to train, but can still be more computationally-efficient than traditional techniques. These nuances highlight the need for AI labs and policymakers to track compute use empirically, to understand how it is evolving, and to project what these trends mean for future demand. In addition to ensuring sufficient access to the right kind of chips, a compute strategy should also prioritise the critical infrastructure and engineering skills needed to manage access and ensure reliability. This is often under-resourced in academia and public research institutions. ■

4. Data

Blend top-down and bottom-up efforts to collect, curate, store, and access data

Similar to compute, data can be viewed as critical *infrastructure* for AI for Science efforts that needs to be developed, maintained, and updated over time. Discussions often focus on identifying new datasets that policymakers and practitioners should create. There is a role for such top-down efforts. In 2012, the Obama Administration launched the Materials Project to map known and predicted materials, such as inorganic crystals, like silicon, that are found in batteries, solar panels, and computer chips. Our recent GNoME effort used this data to predict 2.2 million novel inorganic crystals, including 380,000 that simulations suggest are stable at low temperatures, making them candidates for new materials.

However, it is often difficult to predict in advance what scientific datasets will be most important, and many AI for Science breakthroughs rely on data that emerged more organically, thanks to the efforts of an enterprising individual or small teams. For example, Daniel MacArthur, then a researcher at the Broad Institute, led the development of the gnomAD dataset of genetic variants that our AlphaMissense work subsequently drew on. Similarly, the mathematical proof assistant and programming language Lean was originally developed by the programmer Leonardo de Moura. It is not a dataset, but many AI labs now use it to help train their AI maths models, including our AlphaProof system.

Efforts like gnomAD or Lean highlight how top-down data efforts need to be complemented by better incentives for individuals at all stages of the data pipeline. For example, some data from strategic wet lab experiments is currently discarded, but could be collected and stored, if stable funding was available. Data curation could also be better incentivised. Our AlphaFold models were trained on data from the Protein Data Bank that was particularly high quality because journals require the deposition of protein structures as a precondition for publication, and the PDB's professional data curators developed standards for this data. In genomics, many researchers are also obliged to deposit raw sequencing data in the Sequence Read Archive but inconsistent standards mean that individual datasets often still need to be reprocessed and combined. Some other high-quality datasets go unused altogether, because of restrictive licensing conditions, such as in biodiversity, or because the datasets are not released, such as decades of data from publicly-funded fusion experiments. There can be logical reasons for this, such as a lack of time, funds, somewhere to put the data, or the need for temporary embargo periods for researchers who develop the data. But in aggregate these data access issues pose a key bottleneck to using AI to advance scientific progress. ■

5. Organisational design

Strike the right balance between bottom-up creativity and top-down coordination

A simple heuristic is that academia and industry tend to approach science research at two ends of a spectrum. Academia tends to be more bottom-up, and industry labs tend to be more top-down. In reality, there has long been plenty of space in between, particularly at the most successful labs, such as the golden eras of Bell Labs and Xerox PARC that were renowned for their blue skies research and served as inspiration in the founding of DeepMind. Recently, a new wave of science research institutions has emerged that try to learn from these outlier examples. These organisations differ in their goals, funding models, disciplinary focus, and how they organise their work. But collectively they want to deliver more high-risk, high-reward research, less bureaucracy, and better incentives for scientists. Many have a strong focus on applying AI, such as the UK's Advanced Research and Invention Agency, the Arc Institute, and the growing number of Focused Research Organisations that aim to tackle specific problems in science that are too large for academia and not profitable enough for industry, such as the organisation tasked with expanding the Lean proof assistant that is pivotal to AI maths research.

At their core, these new institutions share a desire to find a better blend of top-down coordination and focus with bottom-up empowerment of scientists. For some organisations, this means focussing on a single specific problem with pre-specified milestones. For others, it means offering more unrestricted funding to principal investigators. Getting this balance right is critical to attracting and retaining research leaders, who must also buy into it if it is to succeed - Demis Hassabis has credited it as the single biggest factor for successfully coordinating cutting-edge research at scale. Striking this balance is also important *within* individual research efforts. In Google DeepMind's case, efforts often pivot between more unstructured 'exploration' phases, where teams search for new ideas, and faster 'exploitation' phases, where they focus on engineering and scaling performance. There is an art to knowing when to switch between these modes and how to adapt the project team accordingly. ■

6. Interdisciplinarity

Approach science as a team, fund neglected roles, and promote a culture of contestability

Many of the hardest scientific problems require progress at the boundaries between fields. However when practitioners are brought together, for example during Covid-19, they often struggle to transition from *multidisciplinary* teams - where they each retain their own disciplinary angle - to genuine *interdisciplinarity*, where they collectively develop shared ideas and methods. This challenge reflects the growing specialisation of scientific knowledge, as well as incentives such as grant funding, that often evaluate practitioners predominantly on their core expertise.

AI for Science efforts are often multidisciplinary by default, but to succeed, they need to become genuinely interdisciplinary. A starting point is to pick a problem that requires each type of expertise, and then provide enough time and focus to cultivate a team dynamic around it. For example, our Ithaca project used AI to restore and attribute damaged ancient Greek inscriptions, which could help practitioners to study the thought, language, and history of past civilizations. To succeed, project co-lead Yannis Assael had to develop an understanding of *epigraphy* - the study of ancient inscribed text. The project's epigraphers, in turn, had to learn how the AI model worked, given the importance of intuition to their work. Cultivating these team dynamics requires the right incentives. Empowering a small, tight-knit team to focus on solving *the problem*, rather than authorship of papers, was key to the AlphaFold 2 breakthrough. This type of focus can be easier to achieve in industry labs, but again highlights the importance of longer-term public research funding that is less tied to publication pressures.

This image shows a restored decree concerning the Acropolis of Athens and dates 485/5 BCE.
(CC BY-SA 3.0, WikiMedia).

To achieve genuine interdisciplinarity, organisations also need to create roles and career paths for individuals who can help blend disciplines. At Google DeepMind, our *research engineers* encourage a positive feedback loop between research and engineering, while our *programme managers* help to cultivate team dynamics within a research effort and create links across them. We also prioritise hiring individuals who enjoy finding and bridging connections between fields, as well as those that are motivated by rapidly upskilling in new areas. To encourage a cross-pollination of ideas, we also encourage scientists and engineers to regularly switch projects. Ultimately, the goal is to create a culture that encourages curiosity, humility and what the economic historian Joel Mokyr has referred to as 'contestability' - where practitioners of all backgrounds feel empowered to present and constructively critique each other's work in open talks and discussion threads. ■

7. Adoption

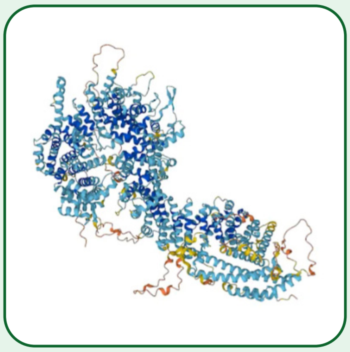
Carefully consider the best access option and spotlight AI models' uncertainties

Many AI for Science models, such as AlphaFold or our weather forecasting work, are *specialised* in the sense that they perform a small number of tasks. But they are also *general* in the sense that a large number of scientists are using them, for everything from understanding diseases to improving fishing programmes. This impact is far from guaranteed. The germ-theory of disease took a long time to diffuse, while the downstream products that scientific breakthroughs could enable, such as novel antibiotics, often lack the right market incentives.

When deciding how to release our models, we try to balance the desire for widespread adoption and validation from scientists with commercial goals and other considerations, such as potential safety risks. We also created a dedicated Impact Accelerator to drive adoption of breakthroughs and encourage socially beneficial applications that may not otherwise occur, including through partnerships with organisations like the Drugs for Neglected Diseases Initiative, and the Global Antibiotic Research & Development Partnership, that have a similar mandate.

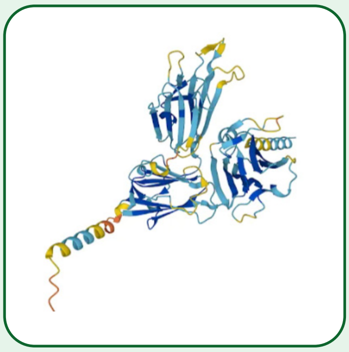
To encourage scientists who could benefit from a new model or dataset to use it, developers need to make it as easy as possible for scientists to use and integrate into their workflows. With this in mind, for AlphaFold 2 we open-sourced the code but also partnered with EMBL-EBI to develop a database where scientists, including those with less computational skills and infrastructure, could search and download from a preexisting set of 200 million protein structures. AlphaFold 3 expanded the model's capabilities, leading to a combinatorial explosion in the number of potential predictions. This created a need for a new interface, the AlphaFold Server, which allows scientists to create structures on-demand. The scientific community has also developed their own AlphaFold tools, such as ColabFold, demonstrating the diversity of needs that exist, as well as the value of nurturing computational skills in the scientific community to address these needs.

AlphaFold predictions referenced in publications



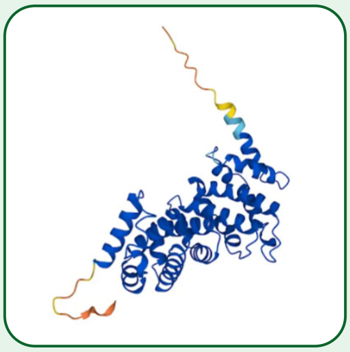
Nuclear pore complex protein Nup205

Part of a large complex that acts as a gateway in and out of the cell nucleus



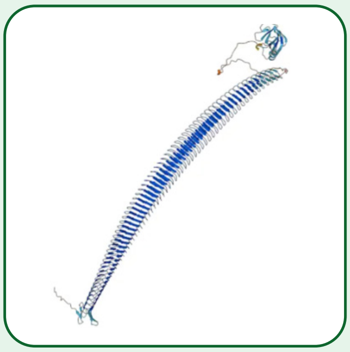
Gametocyte surface protein P45/48

From the maleria parasite; a candidate protein for including in vaccines



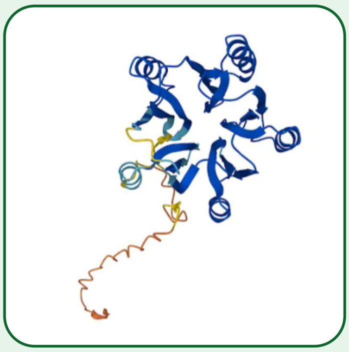
CCR4-NOT transcription complex subunit 9

Regulates an important cellular process (the rate of mRNA degradation)



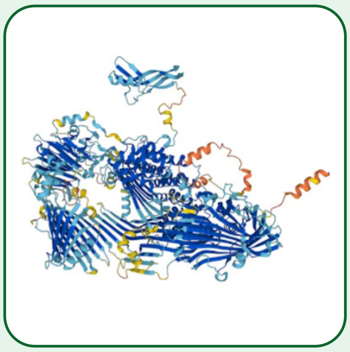
Ice nucleation protein

Bacterial protein that can trigger ice formation at relatively high temperatures, causing frost damage to plants



F20H23.2 protein

Plant protein; represents a potential new structural superfamily unlike anything seen before



Vitellogenin

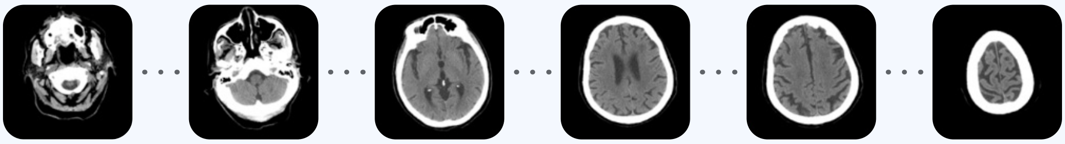
Involved in the immune system of egg-laying animals including honeybees

To date, more than 2 million users from over 190 countries have accessed the AlphaFold Protein Structure Database to view over 7 million structures.

Scientists also need to trust an AI model in order to use it. We expand on the reliability question below, but a useful starting point is to proactively signal how scientists should use a model, as well as its uncertainties. With AlphaFold, following dialogue with scientists the team developed uncertainty metrics that communicated how 'confident' the model was about a given protein structure prediction, supported by intuitive visualisations. We also partnered with EMBL-EBI to develop a training module that offered guidance on how to best use AlphaFold, including how to interpret the confidence metrics, supported by practical examples of how other scientists were using it. Similarly, our Med-Gemini system recently achieved state-of-the-art performance on answering health-related questions. It uses an *uncertainty-guided* approach that responds to a question by generating multiple 'reasoning chains' for how it might answer. It then uses the relative divergence between these initial answers to calculate how uncertain the answer is. Where uncertainty is high, it invokes web search to integrate the latest, up-to-date information. ■

Prompt

Given the image, what are the radiology findings?



Response

FINDINGS: mild cortical atrophy and periventricular white matter changes are noted compatible with microvasculopathy. there is proportionate ventriculomegaly. no bleed, mass effect, midline shift or abnormal intracranial fluid is seen. no evidence of acute intracranial hemorrhage, territorial infarct, abnormal fluid or mass effect.no abnormal extra-axial fluid collection is identified. no definitive evidence of intracranial hemorrhage or large territorial infarct within the limits of ct imaging. **old left basilar ganglia infarct**. no intracranial mass, swelling, hemorrhage, infarction or extra-axial collection is seen. visualized mastoid air cells: well pneumatized. visualized paranasal sinuses: well pneumatized. atherosclerotic calcification of vasculature at the skull base.

Radiologist Written Report

FINDINGS: the fourth ventricle is midline. the remaining ventricular system is within normal limits. there is mild chronic white matter microangiopathic disease and moderate generalized cortical atrophy. no bleed, mass effect, midline shift or abnormal intracranial fluid is seen. bone windows demonstrate an intact skull. the remaining visualized paranasal sinuses and mastoids are clear. chronic white matter microangiopathic disease, mild generalized cortical atrophy, not significantly changed since 2015. no evidence of acute intracranial hemorrhage, territorial infarct, abnormal fluid or mass effect.

Med-Gemini-3D is able to generate reports for CT scans, a far more complex form of imaging than standard X-rays. In this example, Med-Gemini-3D's report has correctly included a pathology (highlighted in green) that was missed in the origional radiologist's report. Note that 'basilar' is a common mis-transcription of 'basal' that Med-Gemini has learned from the training data, though the meaning of the report is unchanged.

8. Partnerships

Aim for early alignment and a clear value exchange

AI for Science efforts require a diversity of expertise that creates a strong need for partnership - both formal and informal - between public and private organisations. These partnerships are needed throughout the project lifecycle, from creating datasets to sharing the research. In particular, AI labs often need scientists to help evaluate an AI model's outputs. For example, recent research highlighted how senior material scientists with rich domain expertise were critical to evaluating whether the new materials predicted by an AI model were likely to be viable. Similarly, our protein design team partnered with research groups from the Francis Crick Institute to run wet lab experiments to test if our AI-designed proteins bound to their target and if this had the desired function, such as preventing SARS-CoV-2 from infecting cells. Our FunSearch method was able to develop a new construction for the Cap Set problem, which renowned mathematician Terence Tao once described as his favourite open question. This was enabled by collaborating with Jordan Ellenberg, a professor of mathematics at the University of Wisconsin–Madison and a noted Cap Set expert. Given the central role played by industry labs in advancing AI capabilities, and the need for rich domain expertise, these public private partnerships will likely become increasingly important to advancing the AI for Science frontier and may require greater investment, such as more funding to support partnerships teams in universities and public research institutions.

Developing partnerships is difficult. When starting discussions, it is important to align early on the overall goal and address potentially thorny questions, such as what rights each party should have over the outputs, whether there should be a publication, whether the model or dataset should be open sourced, and what type of licensing should apply. Differences of opinion are natural and often reflect the incentives facing public and private organisations, which in turn vary greatly, depending on factors such as the maturity of the research or its commercial potential. The most successful partnerships involve a clear value-exchange that draws on the strengths of each organisation. For example, more than 2 million users from over 190 countries have used the AlphaFold Protein Structure Database. This required a close collaboration to pair our AI model with the biocuration expertise and scientific networks of EMBL-EBI. ■

9. Safety & responsibility

Use assessments to explore trade-offs and inspire new types of evaluation methods

Scientists often disagree, sometimes strongly, about the potential benefits and risks that AI models may have on science, and on wider society. Conducting an ethics and safety assessment can help to frame the discussion and enable scientists to decide whether, and how, to develop a given AI model. A starting point is to identify the most important domains of impact, and to specify these domains at the right level of abstraction. There are increasingly sophisticated frameworks to identify and categorise different AI risks, such as enabling mis- and disinformation. But these frameworks rarely consider the potential benefits of AI in the same domain, such as improving access to high-quality information synthesis, or the trade-offs that can occur, for example if you restrict access to an AI model or limit its capabilities. Assessments should also clarify their timescales, the relative certainty of any impact, and the relative importance, or *additionality,* of AI, to achieving it. For example, those worried about AI and climate change often focus on the immediate power needed to train large AI models, while AI proponents often focus on the less immediate, less clear, but potentially much larger downstream benefits to the climate from future AI applications. In carrying out their assessment, AI practitioners should also avoid over-indexing on the model's capabilities, which they will be closer to, and better understand the extent to which third parties will actually use it or be affected by it, which typically requires input from external experts to do well.

Practitioners also need new methods to better evaluate the potential risks and benefits of using AI in science. At present, many AI safety evaluations rely on specifying the types of content that a model should not output, and quantifying the extent to which the model adheres to this policy. These evaluations are useful for certain risks posed by using AI in science, such as generating inaccurate content. But for other risks, such as to biosecurity, the idea that we can reliably specify certain types of scientific knowledge as dangerous in advance has been challenged, because of the dual-use nature of scientific knowledge, but also because such efforts tend to focus on what has caused harm historically, such as viruses from past outbreaks, rather than novel risks. A better approach may be to evaluate the dangerous capabilities of AI models, or the degree to which AI models provide an uplift to humans' dangerous capabilities. In many cases, these capabilities will also be dual-use, such as the ability to help design or execute experimental protocols. The degree to which these AI capabilities point to a risk, or an opportunity, will depend on how potential threat actors are assessed and how access to the model is governed. Beyond safety, evaluating other risks from using AI in science, such as to scientific creativity or reliability (which we discuss below), will require entirely new evaluation methods. Given the difficulty of researching and executing such evaluations, it makes sense to pursue them at the community-level, rather than each lab pursuing siloed efforts. ■

Part C

The risks

The risks

Policy papers, government documents and surveys of scientists regularly cite certain risks from the growing use of AI in science. Three of these risks - to scientific creativity, reliability, and understanding - mainly relate to how science is practised. Two other risks - to equity and the environment - mainly relate to how science represents and affects wider society. The use of AI is often presented exclusively as a risk to these domains, and the domains, such as scientific reliability, or the environment, are often portrayed in stable, somewhat idealised terms, that can overlook the wider challenges that they face. We believe that using AI in science will ultimately benefit each of these five domains, because there are opportunities to mitigate the risks that AI poses, and to use AI to help address wider challenges in these areas, in some cases profoundly. Achieving a beneficial outcome will likely be harder for inequity, which is ingrained into AI and science at multiple levels, from the make-up of the workforce to the data underpinning research, and for scientific creativity, which is highly subjective and so individuals may reasonably disagree about whether a certain outcome is positive. These nuances increase the value of scientists, policymakers and others articulating their expectations for how using AI in science will affect each of these 5 areas. ■

5 risks of using AI to accelerate science

1. Creativity
2. Reliability
3. Understanding
4. Equity
5. Environment

1. Creativity

Will AI lead to less novel, counterintuitive, breakthroughs?

Scientific creativity describes the creation of something new that is useful. In practice, the extent to which a scientist views a new idea, method, or output as creative typically rests on more subjective factors, such as its perceived simplicity, counterintuitiveness, or beauty. Today, scientific creativity is undermined by the relative homogeneity of the scientific workforce, which narrows the diversity of ideas. The pressure on researchers to 'publish or perish' also incentivises 'crowd-following' publications on less risky topics, rather than the kind of deep work, or bridging of concepts across disciplines, that often underpins creative breakthroughs. This may explain why the share of disruptive scientific ideas that cause a field to veer off into a new direction seems to be declining, beyond what may be normally expected, as science expands.

Some scientists worry that using AI may exacerbate these trends, by undermining the more intuitive, unorthodox, and serendipitous approaches of human scientists, such as Galileo's hypothesis that the earth rotates on its axis. This could happen in different ways. One concern is that AI models are trained to *minimise* anomalies in their training data, whereas scientists often *amplify* anomalies by following their intuitions about a perplexing data point. Others worry that AI systems are trained to perform specific tasks, and so relying on them will forgo more serendipitous breakthroughs, such as researchers unexpectedly finding solutions to problems that they weren't studying. At the community level, some worry that if scientists embrace AI *en masse*, it may lead to a gradual homogenisation of outputs, for example if LLMs produce similar suggestions in response to the queries of different scientists. Or if scientists over-focus on disciplines and problems that are best-suited to AI.

Maintaining support for exploratory research and non-AI research could help to mitigate some of these risks. Scientists could also tailor how they use AI so that it boosts rather than detracts from their own creativity, for example by fine-tuning LLMs to suggest more personalised research ideas, or to help scientists better elicit their own ideas, similar to our early efforts to develop AI tutors that could help students to better reflect on a problem, rather than just outputting answers to questions. AI could also enable new types of scientific creativity that may be unlikely to otherwise occur. One type of AI creativity is *interpolation* where AI systems identify novel ideas within their training data, particularly where humans' ability to do this is limited, such as efforts to use AI to detect anomalies in massive datasets from Large Hadron Collider experiments. A second type is *extrapolation*, where AI models generalise to more novel solutions outside their training data, such as the famous move 37 that our AlphaGo system came up with, that stunned human Go experts, or the novel maths proofs and non-obvious constructions that our AlphaProof and AlphaGeometry 2 systems produced. A third type is *invention,* where AI systems come up with an entirely new theory or scientific system, completely removed from its training data, akin to the original development of general relativity, or the creation of complex numbers. AI systems do not currently demonstrate such creativity, but new approaches could potentially unlock this, such as multi-agent systems that are optimised for different goals, like novelty and counterintuitiveness, or AI models that are trained to generate novel scientific problems in order to inspire novel solutions. ■

2. Reliability

Will AI make science less self-correcting?

Reliability describes the ability of scientists to depend upon each others' findings, and trust that they are not due to chance or error. Today, a series of interrelated challenges weaken the reliability of science, including the p-hacking and publication bias which can lead researchers to underreport negative results; a lack of standardisation in how scientists carry out routine scientific tasks; mistakes, for example in how scientists use statistical methods; scientific fraud; and challenges with the peer review process, including a lack of qualified peer reviewers.

Some scientists worry that AI will exacerbate these challenges as some AI research also features bad practices, such as practitioners cherrypicking the evaluations they use to assess their models' performance. AI models, particularly LLMs, are also prone to 'hallucinate' outputs, including scientific citations, that are false or misleading. Others worry that LLMs may lead to a flood of low-quality papers similar to those that 'paper mills' churn out. The community is working on mitigations to these problems, including good practice checklists for researchers to adhere to and different types of AI factuality research, such as training AI models to ground their outputs to trusted sources, or to help verify the outputs of other AI models.

Scientists could also potentially use AI to improve the reliability of the wider research base. For instance, if AI can help to automate aspects of data annotation or experiment design, this could provide much-needed standardisation in these areas. As AI models get better at grounding their outputs to citations, they could also help scientists and policymakers do more systematic reviews of the evidence base, for example in climate change, where groups like the Intergovernmental Panel on Climate Change are already struggling to keep up with the inexorable rise in publications. Practitioners could also use AI to help detect mistaken or fraudulent images, or misleading scientific claims, as seen in the recent trial by the Science journals of an AI image analysis tool. More speculatively, AI could potentially help with aspects of peer review, given that some scientists already use LLMs to help critique their own papers, and to help validate the outputs of AI models, for example in theorem proving. However, there are also reasonable concerns about confidentiality, the ability of AI systems to detect truly novel work, and the need for buy-in from scientists given the consequential role that peer review plays in processes such as grant approvals. ■

3. Understanding

Will AI lead to useful predictions at the expense of deeper scientific understanding?

In a recent Nature survey, scientists cited a reliance on pattern matching at the expense of deeper understanding as the biggest risk from using AI in science. Understanding is not always necessary to discover new scientific phenomena, such as superconductivity, or to develop useful applications, such as drugs. But most scientists view understanding as one of their primary goals, as the deepest form of human knowledge. Concerns about AI undermining scientific understanding include the argument that modern deep learning methods are *atheoretical* and do not incorporate or contribute to theories for the phenomena that they predict. Scientists also worry that AI models are *uninterpretable*, in the sense that they are not based on clear sets of equations and parameters. There is also a concern that any explanation for an AI model's outputs will not be accessible or useful to scientists. Taken together, AI models may provide useful predictions about the structure of a protein, or the weather, but will they be able to help scientists understand why a protein folds a certain way, or how atmospheric dynamics lead to weather shifts?

Concerns about replacing '*real, theoretical science*' with '*low-brow … .computation*' are not new and were levelled at past techniques, such as the Monte Carlo method. Fields that merge engineering and science, such as synthetic biology, have also faced accusations of prioritising useful applications over deeper scientific understanding. Those methods and technologies led to gains in scientific understanding and we are confident that AI will too, even if some of these gains will be hard to predict in advance. First, most AI models are not atheoretical but build on prior knowledge in different ways, such as in the construction of their datasets and evaluations. Some AI models also have interpretable outputs. For example, our FunSearch method outputs computer code that also describes how it arrived at its solution.

Researchers are also working on explainability techniques that could shed light on how AI systems work, such as efforts to identify the 'concepts' that a model learns. Many of these explainability techniques have important limitations, but they have already enabled scientists to extract new scientific hypotheses from AI models. For example, transcription factors are proteins that bind to DNA sequences to activate or repress the expression of a nearby gene. One AI research effort was able to predict the relative contribution of each base in a DNA sequence to the binding of different transcription factors and to explain this result using concepts familiar to biologists. A bigger opportunity may be to learn entirely new concepts based on how AI systems learn. For example, our researchers recently demonstrated that our AlphaZero system learned 'superhuman' knowledge about playing chess, including unconventional moves and strategies, and used another AI system to extract these concepts and teach them to human chess experts.

Even without explainability techniques, AI will improve scientific understanding simply by opening up new research directions that would otherwise be prohibitive. For example, by unlocking the ability to generate a huge number of synthetic protein structures, AlphaFold enabled scientists to search *across protein structures,* rather than just across protein sequences. One group used this approach to discover an ancient member of the Cas13 protein family that offers promise for editing RNA, including to help diagnose and treat diseases. This discovery also challenged previous assumptions about how Cas13 evolved. Conversely, efforts to modify the AlphaFold model architecture to incorporate more prior knowledge led to worse performance. This highlights the trade-off that can occur between accuracy and interpretability, but also how AI systems could advance scientific understanding not *in spite of* their opacity, but *because* of it, as this opacity can stem from their ability to operate in high-dimensional spaces that may be uninterpretable to humans, but necessary to making scientific breakthroughs. ■

4. Equity

Will AI make science less representative, and useful, to marginalised groups?

Inequity is starkly visible in the scientific workforce, in the questions they study, in the data and models they develop, and in the benefits and harms that result. These inequities are related and can compound over time. For example, a small number of labs and individuals in higher-income cities account for a disproportionate share of scientific outputs. Studies to identify genetic variants associated with disease rely heavily on data from European ancestry groups, while the neglected tropical diseases that disproportionately affect poor countries receive relatively little research funding. In agriculture, crop innovations focus on pests that are most common in high-income countries, and are then inappropriately used on different pests in lower-income countries, hurting yields. Despite positive trends, women account for just 33% of scientists and have long been underrepresented in clinical trials, particularly women of colour.

Observers worry that the growing use of AI in science could exacerbate these inequities. AI and computer science workforces are less representative, in terms of gender, ethnicity and the location of leading labs, than many other scientific disciplines and so AI's growing use could hurt broader representation in science. As a data-driven technology, AI also risks inheriting and entrenching the biases found in scientific datasets.

There are also opportunities to use AI to reduce inequities in science, albeit not in lieu of more systemic change. If AI models are provided via low-cost servers or databases, they could make it easier and cheaper for scientists, including those from underrepresented groups, to study traditionally neglected problems, similar to how releasing more satellite data led to more research from underrepresented communities. By ingesting more data, AI models may also be able to learn more universal patterns about the complex systems that scientists study, making these models more robust and less prone to biases. For example, owing to their non-representative data, studies that identify genetic variants associated with disease can pick up *confounding,* rather than *causal* variants. Conversely, some early attempts to train AI models on larger datasets of protein structures and genetic variants, including data across species, perform better at predicting individuals at the greatest risk for disease, with fewer discrepancies across population groups. Ultimately, however, improving equity will require long-term efforts, such as the H3Africa initiative in genomics and the Deep Learning Indaba initiative for AI, that aim to build up scientific infrastructure, communities, and education where it is most lacking. ■

5. The environment

Will AI hurt or help efforts to achieve NetZero?

Given their desire to understand the natural world, many scientists have long been active in efforts to protect the environment, from providing early evidence about climate change to developing photovoltaic cells. In recent years, a growing number of scientists have voiced concerns about the potential impact of AI on the environment and developed methodologies to try to quantify these impacts. Most concerns focus on the potential impact of training and using LLMs on greenhouse gas emissions, as well as related concerns, such as about the water needed to cool data centres. One way to think about these effects is the *life cycle approach*, which captures both direct and indirect effects. Direct effects include the emissions from building and powering the data centres and devices that AI models are trained and run on. There is no comprehensive estimate for *all* direct emissions from AI. However, a 2021 estimate suggested that cloud and hyperscale data centres, where many large AI models are trained and deployed, accounted for just 0.1-0.2% of global emissions.

As the size of LLMs continues to grow, observers have cautioned that these figures may increase, potentially significantly. However, many users of LLMs, including scientists, will be able to fine-tune them, or use their predictions, at a relatively low compute cost, rather than training them from scratch. Efforts are also underway to make LLMs more efficient, and the history of digital technology suggests that sizable gains are possible, not least due to the commercial pressures to deliver faster and cheaper AI models. In some instances, the emissions from AI models will be lower than other approaches. For example, our internal analysis suggests that determining the structures of a small number (<10) of proteins experimentally uses roughly the same energy as a full training run of AlphaFold 2. These results need to be interpreted carefully, as AI simulations rely on, and inform, physical experiments, rather than substituting for them. But they also show how AI could enable a larger amount of scientific activity at a lower average energy cost.

Crucially, the *direct* effects of AI on emissions, whether positive or negative, will likely be minor compared to the *indirect* effects that AI-enabled applications have on emissions. Using AI in science opens up three major opportunities to reduce emissions. First, progress at the nexus between AI, maths and computer science could dramatically improve the efficiency of the Internet, from designing more efficient chips to finding more efficient algorithms for routine tasks. As a growing share of the economy moves online, this should help to offset emissions across these sectors. AI could accelerate the development and use of renewable energy, for example by designing new materials, such as for batteries or solar panels, by optimising how the grid operates and how it integrates renewables, and via more transformative but uncertain opportunities like fusion. Finally, the world is already getting warmer and AI could help to better prepare for extreme weather events. For example, our weather forecasting model recently correctly predicted, seven days in advance, that the deadly Hurricane Beryl would make 'landfall' in Texas. Non-AI models had originally predicted landfall in Mexico before correcting their prediction to Texas three days before it occurred. ■

Part D

The policy response

The policy response

Given the importance of scientific progress to almost every major economic, environmental and security goal, it follows that science, and the potential for AI to accelerate it, should be a top priority for any government. What should a new AI *for Science* policy agenda look like? Policymakers can start by implementing the many good science and innovation policy ideas that already exist and which make even more sense in an era of AI-enabled science. For example, AI will improve the return on science research funding and so it provides a strong rationale to invest more in it and to trial new ideas to speed up and experiment with how this funding is allocated. On compute, governments could implement the idea set out in the UK's Independent Review to empower a dedicated body to continually assess and advise governments on potential investments. To support AI for Science startups, policymakers can improve their spin-out policies and support well-run start-up incubators and fellowships.

But ambitious new policies are also needed to capitalise on the AI for Science opportunity. We share **four ideas** below. They are intended to be widely applicable, though the precise details would need to be tailored to the specific context of a country, taking into account national priorities, unique strengths and the institutional landscape. ■

1. Define the 'Hilbert Problems' for AI in Science

Scientific progress rests on picking the right problems. In 1900, the German mathematician David Hilbert published 23 unsolved problems that proved hugely influential for the subsequent direction of 20th century mathematics. As part of upcoming international events such as the AI Action Summit in Paris, policymakers, AI labs and science funders could launch a public call for scientists and technologists to identify the most important AI-shaped scientific problems, backed by a major new global fund to drive progress on them. Submissions should specify why the problem is important, why it is suited to modern AI systems, why it may be otherwise neglected, the data bottlenecks that exist, and how near-term technical progress could be evaluated.

The top ideas could form the basis of new scientific competitions, where scientists compete to solve these problems with AI, supported by new datasets, evaluation methods and competitive benchmarks. These could build on the recent flurry of competitions that have emerged to evaluate the scientific capabilities of AI models, and include a new AI for Science Olympiad to attract exceptional young talent from across the world to the field. Beyond its direct impacts, the AI for Science 'Hilbert Problems' initiative could provide a welcome focal point for international scientific collaboration and funding, and inspire a new generation of interdisciplinary scientists to identify and pursue AI-shaped problems. ■

2. Make the world readable to scientists

Most scientific data is uncollected, partial, uncurated or inaccessible, making it unavailable to train AI models. There is no single policy response to what is far from a uniform challenge. Policymakers and funders will need to blend a small number of top-down initiatives with support to scale up promising grassroots efforts. A new international network of *AI for Science Data Observatories* should be set up to help address these goals. These Observatories could be given long-term backing and tasked with running rapid AI for Science 'data stocktakes', where expert teams map the state of data in priority disciplines and application areas. Stocktakes could identify existing datasets, such as the Sequence Read Archive, whose quality could be further improved, as well as untapped or underutilised datasets, such as the decades of experimental fusion data that is currently unavailable to scientists or leading biodiversity datasets that are subject to restrictive licensing conditions. The stocktakes could also include new 'data wish lists'. For example, our internal analysis suggests that less than 7% of papers in key environmental research domains use AI. We recently funded Climate Change AI to identify datasets which, if available or improved, could remove some of the bottlenecks to higher AI use. To ensure this analysis leads to action, policymakers should designate and empower organisations to be accountable for addressing the results of the data stocktakes.

The observatories could also scope the creation of new databases, including ensuring that adequate consideration is given to their long-term storage, maintenance, and incentives. This could include new databases to securely store the results of strategic wet lab experiments that are currently discarded, complemented by making the deposition of these experimental results a requirement for public research funding. Or digitising more public archives, following the example of a recent UK government and Natural History Museum collaboration to digitise their natural science collections, which includes more than 137 million items, from butterflies to legumes, across a 4.6 billion-year history. Policymakers can also empower scientists to use LLMs to create and improve their own datasets, by ensuring that publicly-funded research is open by default, where possible, building on recent examples from the UK, US and Japan, including mandates to release research via pre-print servers. Policymakers could seek to co-fund the most ambitious dataset initiatives with industry and philanthropy. ■

---

3. Teach AI as the next scientific instrument

Over the past half century, as the number of scientific technologies has grown, so has most scientists' distance from them. Many technologies are the products of science, but an ever smaller share of scientists are trained in how to develop and use them effectively. The pressing near-term need is to fund and incentivise mass uptake of shorter, more tactical AI training programmes and fellowships, for existing scientists and research leaders. Policymakers can incentivise these efforts by setting a clear goal that every postgraduate science student should be able to access introductory courses on using AI in science, including on the most important tools in their domain, in the same way as basic statistics is often taught today. The type and depth of training needed will depend on an individual's discipline and profile, and could range from basic introductory courses about how to reliably use LLMs for everyday research tasks, through to more advanced courses on how to fine-tune AI models on scientific data, as well as how to address more complex challenges, such as evaluating whether the data they used to test a model's performance has intentionally or unintentionally 'leaked' into the data used to train it.

These programmes could build on established examples such as the University of Cambridge's Accelerate Programme that provides structured training in AI to PhD and postdoctoral researchers, or the short courses that The Carpentries offer on the programming, data, and computational skills needed to do research.

Policymakers also need to move quickly to put in place longer-term programmes to ensure that the next generation of scientists has the skills they need. This means mainstreaming and deepening AI training and skills development in science education at all levels. Secondary school science students will need early exposure to the impact of AI while university students will need access to new types of interdisciplinary AI science degrees, such as the pan-African AI for Science Masters programme that we partnered with the African Institute for Mathematical Sciences to develop. Dedicated scholarships could also help. For example, the UK's BIG Scholarships programme provides outstanding opportunities to high school students, with a focus on those from underrepresented groups who have excelled in International Science Olympiads and want to continue their study in leading science hubs but lack the funds to do so. ■

4. Build evidence and experiment with new ways of organising science

Scientists' use of AI is growing exponentially, but policymakers have little evidence about who is doing it best, how they are doing it, and the hurdles that are inhibiting others. This evidence gap is an impediment to identifying the best AI for Science policy ideas and targeting them effectively. Historically, answers to such questions often come from fields such as economics or innovation studies, but the results can take years to arrive. We are using citation data analysis, interviews, and community engagement to understand how scientists are using our AI models. Governments are also investing in these *metascience* capabilities to improve how they fund, share and evaluate science research. Building on this momentum, scientists could be tasked with a mission to rapidly assess foundational policy questions, including: where is the most impactful AI for Science research occurring and what types of organisations, talent, datasets, and evaluations are enabling it? To what extent are scientists using and fine-tuning LLMs vs more specialised AI models, and how are they accessing these models? To what extent is AI *actually* benefiting or harming scientific creativity, reliability, the environment, or other domains? How is AI affecting a scientist's perception of their job and what skills, knowledge gaps, or other barriers are preventing their broader use of AI?

Beyond informing robust policy responses, this evidence base will arm policymakers with the foresight they need to anticipate how AI will transform science and society, similar to the foresight they are developing for AI safety risks through the growing network of AI Safety Institutes. The evidence will also highlight opportunities to reimagine the incentives and institutions needed for science in the age of AI. In particular, scientists and policymakers have only explored a small fraction of the possible approaches to organising and executing science research. The rise of AI provides a welcome forcing function to experiment with new types of institutions, from those with more freedom to pursue high-risk, high-reward research, to Focused Research Organisations aimed at addressing specific bottlenecks. And from new types of interdisciplinary AI science institutes in priority domains such as climate or food security, to completely novel institutions that we are yet to imagine. Those who experiment faster will stand to benefit the most from a new golden age of discovery. ■